Essays on optimal experimentation

Willems, T.

Chapter 3

Actively Learning by Pricing: A Model of an Experimenting Seller

Taniyama was not a very careful person as a mathematician. He made a lot of mistakes, but he made mistakes in a good direction, so eventually he got the right answers. I tried to imitate him, but I found out that it is very difficult to make good mistakes.

Goro Shimura on Yutaka Taniyama

3.1 Introduction

This paper investigates the extent to which the active learning motive of a seller who faces uncertainty on the slope of his demand curve can reconcile the volatile, discrete pattern followed by individual prices, with the sluggishness observed in the aggregate price level.

With the arrival of studies analyzing higher-frequency data on prices at the micro level (such as Bils and Klenow (2004)), the type of questions economists have tried to answer has undergone a remarkable twist: while many early contributions tried to explain why firms are *so reluctant* to change their prices (see *e.g.* Greenwald and Stiglitz (1989) and the fixed/sticky price literature), the micro data have shown us that the real puzzle may rather be why firms reprice their products *so often* (Eden and Jaremski, 2010). In particular, Kehoe and Midrigan (2010) and Eichenbaum, Jaimovich and Rebelo (2011) report that posted retail prices tend to change once every two to three weeks. Figure 1, which shows some prices posted by a large US retailer at the weekly frequency, nicely illustrates this.

However, next to the aforementioned studies that document the flexibility of individual prices, there is also evidence that the aggregate price level responds sluggishly to shocks (*cf.* Christiano, Eichenbaum and Evans (1999)).

Lately, the profession has made considerable progress in bringing these two (seemingly conflicting) observations together. On the one hand, papers like Gertler and Leahy (2008), Nakamura and Steinsson (2010) and Kehoe and Midrigan (2010) manage to replicate these dimensions of the data by refining the standard menu cost model. On the other hand, others have pointed out that informational imperfections may also play a role in this: Nimark (2008) for example shows that higher order expectations can reconcile individual price flexibility with aggregate price stickiness, while Maćkowiak and Wiederholt (2009) and Matějka (2010a,b) can match these observations by modeling attention as a scarce resource (following Sims (2003)).

However, high-frequency price data have also shown that individual prices are characterized by at least one other interesting feature: next to their flexibility, they simultaneously show a rigidity. This rigidity however does not take the form of price stickiness in the traditional sense, but rather implies that prices tend to bounce back and forth between only a few rigid values - thereby showing a lot of discreteness (see Figure 1). Matching this dimension of the data with "continuous" models (*i.e.* models without an exogenous preference towards discreteness), turns out to be more challenging. In fact, from the aforementioned papers that manage to reconcile individual price flexibility with aggregate price stickiness, only the ones by Matějka (2010a,b) yield discrete pricing patterns from a continuous setup.[1]

This paper therefore develops a different explanation for the aforementioned micro-macro conflict in pricing behavior - one that is also able to generate individual price series that show a lot of discreteness. It starts by relaxing the strong (but standard) assumption that firms can observe their entire demand curve. Instead, firms are uncertain on the slope of the demand curve for their product.

[1] Gabaix (2011) also obtains discreteness, but he does not analyze his model's ability to reconcile individual price flexibility with aggregate price stickiness.

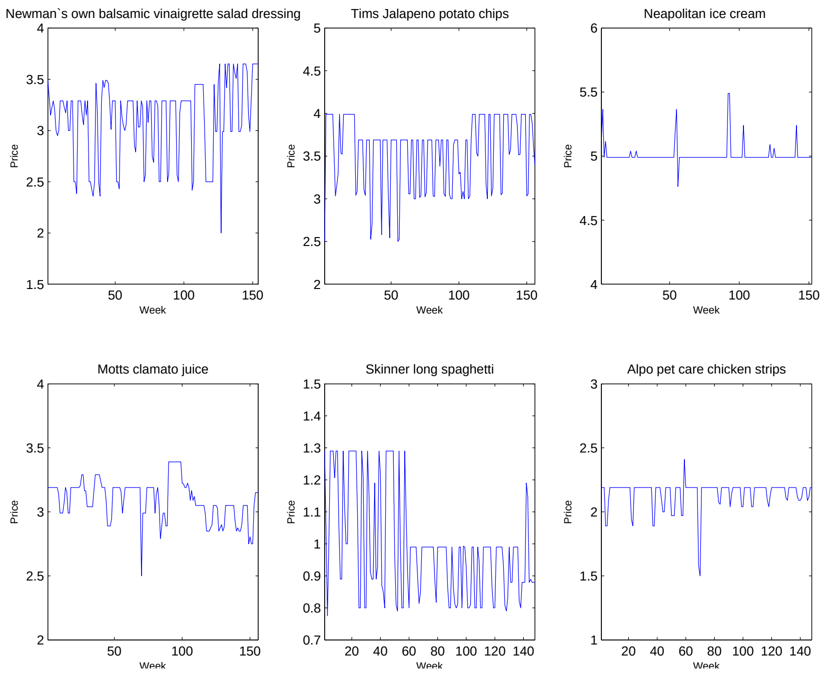

Figure 3.1: Prices posted by a large US retailer for various goods (source: Eichenbaum, Jaimovich and Rebelo (2008)).

Given this uncertainty, firms try to learn the true value of this parameter - thereby endogenously generating an estimate of their demand curve. The way in which this learning process is modeled, however differs from the passive approach that Sargent (1999) for example used to analyze the Fed's learning process on the Phillips curve. Under that strategy, the estimation stage is separated from the control stage as a result of which the resulting learning rule is not optimal.

The present paper on the other hand does optimize the learning rule by taking the links between estimation and control into account. It does so by letting firms learn *actively* about their demand curve. Under active learning, a seller realizes that he is learning from self-generated observations (*i.e.* that he is like an econometrician who has the luxury of being able to generate his own data points). Consequently, he recognizes that there is a link between the actions he takes, and the amount of information that is generated on the parameters he is trying to learn about. As a result, this seller becomes willing to take somewhat more extreme actions that are suboptimal in the short run, because he realizes that these actions generate information - thereby enabling him to achieve higher pay-offs in the future.

Through this channel, the present paper provides a rationale for the occurrence of sales. This phenomenon has lately received a lot of attention from macroeconomists, as their prevalence has raised the question whether sales are important for the dynamics of the aggregate price level. Other recent papers have explained the existence of sales by referring to the formation of implicit contracts between buyers and sellers (Nakamura and Steinsson, 2011), the presence of consumers with different price elasticities (Guimarães and Sheedy, 2011), information processing constraints (Matějka, 2010a,b) or by assuming that price adjustment costs for a sale are lower than those accompanying a regular price change (Kehoe and Midrigan, 2010). The (complementary) motive proposed in this paper is that sales result from the desire of firms to obtain more information on the slope of their demand curve (which enables them to achieve higher profits in the future).

If this learning motive indeed plays a role for putting items on sale in reality, one would expect that the frequency of sales is increasing in the uncertainty on demand. And as documented by Pashigian (1988), this actually is the case. Next to that, there is also more direct evidence that sellers experiment with their price. Einav *et al.* (2011) for example find that experimentation is extremely common on eBay: of the 100 million listings appearing on eBay on a given day, more than half will reappear again as a separate listing, often with modified sale parameters (such as price). Consequently, they state: "Particularly interesting from an economic viewpoint is the possibility that sellers engage in active experimentation to improve their business practices" (p.25). Campbell and Eden (2010) in turn analyze scanner data from US supermarkets and report that grocers deliberately seem to select extreme prices which they then quickly abandon. This suggests to them that "sellers extensively experiment with their prices" (p.1). Their hypothesis is confirmed by survey evidence from the marketing literature: in a questionnaire held among 32 large US retailers, 90 percent of them say to price-experiment (Gaur and Fisher, 2005).[2] The fact that an increasing number of sellers has started to

[2] In September 2000, Amazon.com actually made the headlines with its experimental behavior: users found out that the same DVDs were simultaneously offered to different buyers at different prices. Amazon stated that these differences were due to price experimentation, but critics suggested that they were based on consumer profiling. Since the latter is perceived as being unfair, Amazon promised to refrain from cross-sectional price experimentation in the future - leaving only the time-series dimension to do so. See "Have They Got a Deal For You" in *The Washington Post* of June 19, 2005 for more on this case.

use pricing optimization software (algorithms that tend to take the experimentation motive explicitly into account - very much like in this paper's model), only adds to this.[3]

More generally, the marketing literature tells us that retailers are well aware of the challenges associated with optimal price setting in the face of demand uncertainty. Retailers themselves for example name "measuring the impact of pricing decisions" as their top operational business challenge,[4] in response to which the revenue management literature has produced a large number of papers that looks at exactly this problem (see Araman and Caldentey (2009) for an example).

In this paper, I show that the price paths resulting from the active learning process contain a lot of discreteness: despite the fact that the model consists of continuous equations, prices bounce back and forth between only a few rigid values, just as in the data (*cf.* Figure 1). The reason is that the price setting rules under active learning take the form of a step function: there is a low price platform (associated with low demand) and a high one (for high demand), separated by a sharp kink (which serves to avoid the posting of uninformative prices). In simulations, the resulting price paths look very much like those generated by Matějka's (2010a,b) rationally inattentive seller and consumer models, as they also manage to produce discrete pricing patterns without having an exogenous preference towards it.[5]

In sum, this paper shows that modeling the seller as an optimal experimenter is able to replicate three aspects of the data: the experimentation motive generates volatile pricing patterns that show a lot of discreteness, while the fact that there still is some learning going on makes the aggregate price level respond sluggishly to shocks. Hence, once one considers a model that optimizes the learning process itself, it turns out that learning is not only able to match the aggregate dimension of the data (which is sort of well-known by now), but that it is also capable of

[3] According to the 2004 "Retail Horizons"-survey conducted among 300 US retailers, 36 percent of the responding retailers used pricing optimization software (notable users include Walmart and 7-Eleven, as well as all major hotel and airline companies). In their 2002 survey, this number was only 12 percent. If this trend (along with the increasing importance of online selling) continues, the active learning process of sellers is likely to become even more important in the future as both developments greatly facilitate the practical implementation of optimal price experimentation.

[4] See the RSR-report "Getting Back to Good: Retail Pricing 2010".

[5] As I will explain later on, the discreteness that arises in the present paper is not exact, but looks very much like it. In the rational inattention papers of Matějka (2010a,b) on the other hand, the discreteness is exact. See Matějka and Sims (2010) for a general discussion of discreteness in models with rationally inattentive agents.

reproducing important micro-elements of it. Do note that this result is *not* driven by some form of irrationality: sellers in the model are just responding optimally to the fact that they face a demand curve with an unobserved slope that varies over time.

The remainder of this paper is structured as follows. First, Section 2 provides a short, non-exhaustive overview on the use of active learning in economics (see Kendrick, Amman and Tucci (2011) for a broader survey). Section 3 then describes the model, after which Section 4 explains the solution method. Section 5 contains the calibration and model results, after which a discussion (Section 6) and conclusion (Section 7) complete the paper.

3.2 Relation to the literature

Active learning was first introduced in the economics literature by MacRae (1972) and Prescott (1972).[6] They both considered general optimal control problems in the presence of an experimentation motive. Subsequently, active learning has for example been applied to experimental drug consumption (*cf.* Grossman, Kihlstrom and Mirman (1977)), investment and growth under uncertainty (Bertocchi and Spagat, 1998) and to optimal monetary policy (see Wieland (2000b; 2006), Ellison and Valla (2001) and Svensson and Williams (2007)).

The first application of the active learning-concept to a monopolist who tries to learn his demand curve can be found in Rothschild (1974). He framed the learning process as a two-armed bandit problem and noted that this problem contains a trade-off between long and short run profit maximization (the former calling for price experimentation, while the latter goal is achieved by posting the price that present information indicates is most profitable). Subsequently, several papers have derived important analytical results in this context: Trefler (1993) for example establishes results that determine the direction of experimentation, while Keller and Rady (1999) investigate how the experimentation incentives vary with the discount rate and the persistence of the demand curve.[7]

[6] The concept of active learning is also referred to as "optimal experimentation" (Wieland (2000a), Cosimano (2008)) or "optimal learning" (Wieland, 2000b). It originated in the engineering literature, where it is known as "dual control".

[7] Contemporaneous work by Bachmann and Moscarini (2011) also analyzes a model in which sellers have an experimentation motive, but they mainly use their model to show that macroeconomic fluctuations can generate fluctuations in uncertainty (as measured by price dispersion). I

Following advances in computational power, the present paper is able to take a more practical approach that visualizes the experimentation motive. In this sense, it also relates to Kiefer (1989), who solved Rothschild's (1974) two-armed bandit version of the problem numerically and analyzed how the properties of the value and policy functions varied with the discount factor. In contrast to the present paper, he however did not analyze what this implies for the paths followed by individual and aggregate prices, nor did he establish the link with sales (a phenomenon that did not seem to be of much interest to macroeconomists at the time).

In line with most of the aforementioned contributions, this paper will assume that firms are using their pricing strategy only to learn about the slope of their demand curve. In sharp contrast, there is assumed to be no learning motive on the intercept term.[8] Balvers and Cosimano (1990), which is another important predecessor to the present paper, consider the other polar case in which firms are only trying to learn the intercept of their demand curve in an active manner. They show that this provides firms with an incentive to mute their price adjustments in response to changes in demand. Under their specification, active learning thus implies *the absence of price changes*. Consequently, their model can replicate the delayed response of the aggregate price level in response to shifts in demand, but is unable to match the recently uncovered observation that individual prices change so frequently. This paper's specification on the other hand, in which firms are learning the slope of their demand curve in an active way, turns out to be able to reproduce both dimensions of the data (see Section 5).

3.3 Model

Consider a price-setting seller i who operates in a monopolistically competitive market.[9] He sells his product i in period t at price $p_{i,t}$ and serves whatever

will return to the relevance of Bachmann and Moscarini's work for the present paper in Section 6.

[8] Due to the dimensionality curse associated with the solution method, solving the full problem in which firms are learning on both slope and intercept is very difficult. See Section 6 for a further discussion of this issue and an idea on how the addition of a learning motive for the intercept would affect the results.

[9] The retailer I consider sells his product at only one location. If he would posses multiple selling points, he could in principle also use the cross-sectional dimension to experiment. However, if market power differs between different locations (for example due to the fact that store A does

quantity is demanded at that price. There are no price adjustment costs, so $p_{i,t}$ is set on a period-by-period basis in a fully flexible way. The seller faces a marginal production cost that is equal to c, which is assumed to be known and constant over time. Indicating period t demand for his product by $q_{i,t}$ and using δ to denote the discount factor, his profit maximization problem reads:

$$\max_{p_{i,t}} \mathbb{E}_0 \left\{ \sum_{t=0}^{\infty} \delta^t \left[p_{i,t} - c \right] q_{i,t} \right\} \tag{3.1}$$

He believes that demand for his product follows a linear specification.[10] In particular:

$$\begin{aligned} q_{i,t} &= \alpha_t - \beta_{i,t} \left[p_{i,t} - \overline{p}_t \right] + \gamma q_{i,t-1} \\ &\equiv \alpha_t - \beta_{i,t} r_{i,t} + \gamma q_{i,t-1} \end{aligned} \tag{3.2}$$

Here, $\overline{p}_t$ represents the aggregate price level. Its value lies beyond the control of each individual firm, but can be observed without error. Furthermore, $r_{i,t} \equiv p_{i,t} - \overline{p}_t$ is the firm's relative price, α_t is the aggregate demand component that is common to all firms, and $\beta_{i,t}$ is the slope of the demand curve for the good under consideration. This slope expresses the degree of product differentiation perceived by consumers (do consumers see any close substitutes to variety i?), as a result of which it operates over firm i's *relative* price. Following evidence from the marketing literature (*cf.* Gordon, Goldfarb and Li (2011) and the references therein), this parameter will be time-varying.

There are numerous reasons why the perceived degree of product substitutability can vary over time: it may be due to aggregate conditions (consumers becom-

have a nearby competitor, while store B does not), it is not clear whether there is much to learn for A from B and vice versa. Moreover, supermarkets in some countries (such as the UK) have made agreements with authorities to follow a "national pricing policy", which implies that the same price is charged at all selling points. In practice, retail chains also do not always share information among all sellers, for example due to the fact that different stores have different owners (franchising). Finally, experimenting cross-sectionally is difficult in an online environment (recall footnote 2).

[10] Whether this belief is correct or not is not important (in this paper it will be correct though, so that there is no misspecification bias). The linearity assumption is not essential either (*cf.* Keller and Rady (1999)): the only thing that matters is that the seller estimates a regression equation in his learning step that contains a price sensitivity term which operates over the *relative* price he posts. The linear demand curve (2) satisfies this requirement, as does the CES demand curve (the other popular demand specification).

ing more price sensitive in recessions), store-specific ones (a competitor opening a nearby store, increasing the price sensitivity of shoppers), while brand-specific conditions may also play a role (think of a successful advertising campaign making consumers less price sensitive towards your product, or bad press establishing the reverse). In this paper, all of these factors are captured by the stochastic process for $\beta_{i,t}$ (to be specified below).

As both the industrial organization and marketing literature have generated substantial evidence that consumer behavior is subject to some form of brand loyalty (*cf.* Klemperer (1995), Chintagunta, Kyriazidou and Perktold (2001) and the references therein), I allow for the formation of good-specific habits via the parameter γ.[11] The value of this parameter is assumed to be constant and known to the seller. It should however be stressed that the experimentation motive would still be present without any form of habit formation, so from that point of view its inclusion is not essential (see *e.g.* Kiefer (1989), Wieland (2000a) and Bachmann and Moscarini (2011) who analyze active learning problems without a lagged dependent variable). As will be explained in Section 5 of this paper, the presence of good-specific habits does play a role in generating individual price volatility.

The crucial difference with the standard model is the seller's information set. Whereas the standard model makes the rather strong assumption that sellers can perfectly observe their entire demand curve, this paper relaxes this assumption and keeps the intercept α_t and slope $\beta_{i,t}$ unobserved to the seller. Apart from the (p, q)-point he is producing at every period, the seller thus faces uncertainty on the demand curve for his product. Instead, the seller only knows the stochastic processes that α_t and $\beta_{i,t}$ follow.

Although in practice both the intercept and slope of the demand curve are estimated to be persistent, this paper makes the simplifying assumption that aggregate demand α_t follows an i.i.d.-process with known mean $\overline{\alpha}$, hence:

$$\alpha_t = \overline{\alpha} + \mu_t, \text{ with } \mu_t \sim N(0, \sigma_\mu^2) \tag{3.3}$$

Consequently, there is no learning motive for the intercept term (since it has no

[11] To keep the analysis clear, I abstract from the time-consistency issues that are central to Nakamura and Steinsson (2011). As a result, the commitment-channel they emphasize is not at work in the present paper.

persistence its expected value is always equal to $\overline{\alpha}$). This keeps the number of state variables to a minimum, thereby facilitating the process of finding a numerical solution to the model (see Section 6 for a more detailed discussion of this issue).

The unobserved process for $\beta_{i,t}$ is assumed to be persistent. As a result, sellers in the model do have an incentive to learn more about the actual value of that parameter. I assume that $\beta_{i,t}$ follows an $AR(1)$ around known mean $\overline{\beta}$ (where the latter is common to all sellers, representing aggregate economic conditions):

$$\beta_{i,t} = \left(1-\rho_\beta\right)\overline{\beta} + \rho_\beta \beta_{i,t-1} + \eta_{i,t}, \text{ with } \eta_{i,t} \sim N(0, \sigma_\eta^2) \tag{3.4}$$

The timing and informational structure of the model are as follows: at the beginning of period t, firms have prior beliefs regarding the unobserved parameter they are trying to learn about, $\beta_{i,t}$. These prior beliefs can be described by a conditional normal distribution:

$$p\left(\beta_{i,t}|\Xi_{i,t-1}\right) = \mathcal{N}\left(b_{i,t|t-1}, \Sigma^b_{i,t|t-1}\right)$$

Here, notation is such that $\mathbb{E}_{t-1}\left[\beta_{i,t}\right] \equiv b_{i,t|t-1}$ and $var_{t|t-1}\left[\beta_{i,t}\right] \equiv \Sigma^b_{i,t|t-1}$. From this expression one can see that $\Xi_{i,t-1}$, the vector of state variables describing beliefs, contains two elements: $b_{i,t|t-1}$ and $\Sigma^b_{i,t|t-1}$.

Given these beliefs, firm i sets its relative price $r_{i,t}$, after which all shocks materialize and the firm gets to observe realized demand $q_{i,t}$. With this new information, the firm then updates its beliefs via Bayes' rule, which states that:

$$p\left(\beta_{i,t}|\Xi_{i,t}\right) = \frac{p\left(q_{i,t}|\beta_{i,t}, r_{i,t}, \Xi_{i,t-1}\right) p\left(\beta_{i,t}|\Xi_{i,t-1}\right)}{p\left(q_{i,t}|r_{i,t}, \Xi_{i,t-1}\right)}$$

Hence, we can write the Bayesian updating equations as (*cf.* Zellner (1971)):

$$b_{i,t|t} = b_{i,t|t-1} - \frac{\Sigma^b_{i,t|t-1} r_{i,t}}{r^2_{i,t}\Sigma^b_{i,t|t-1} + \sigma^2_\mu}\left(q_{i,t} - \gamma q_{i,t-1} - \overline{\alpha} + b_{i,t|t-1} r_{i,t}\right) \tag{3.5}$$

$$\Sigma^b_{i,t|t} = \Sigma^b_{i,t|t-1} - \frac{\left(\Sigma^b_{i,t|t-1}\right)^2 r^2_{i,t}}{r^2_{i,t}\Sigma^b_{i,t|t-1} + \sigma^2_\mu} \tag{3.6}$$

Finally, in order to translate the posterior at the end of period t into the prior for the beginning of period $t+1$, we have that:

$$b_{i,t+1|t} = \left(1-\rho_{\beta}\right)\bar{b} + \rho_{\beta} b_{i,t|t} \tag{3.7}$$

$$\Sigma^{b}_{i,t+1|t} = \rho^{2}_{\beta}\Sigma^{b}_{i,t|t} + \sigma^{2}_{\eta}, \tag{3.8}$$

Apart from the fact that these updating equations are non-linear functions of $q_{i,t}$ and $r_{i,t}$, one should also note that they establish a link between different periods since actions taken in the past affect (the quality of) future beliefs. This adds a dynamic dimension to the firm's problem.

Furthermore, active learning endogenizes the model's information structure. In particular, the firm's speed of learning, which is given by the Kalman gain-like expression $\Sigma^{b}_{i,t|t-1} r_{i,t} / \left(r^{2}_{i,t}\Sigma^{b}_{i,t|t-1} + \sigma^{2}_{\mu}\right)$ in equation (5), can be affected by the firm's decisions. More precisely, it is a function of the relative price $r_{i,t}$ firm i sets. The reason is that by setting its relative price in an "informative way", a firm can generate more information on the system it is trying to learn about - thereby facilitating the learning process. This contrasts with standard signal extraction problems, such as the one in Lucas (1973), in which the speed of learning is simply a function of the exogenously given signal-to-noise ratio.

In the present setting on the other hand, sellers can choose to receive more precise signals by letting the control $r_{i,t}$ take on more extreme values.[12] After all, if one estimates a regression that has the form of equation (2) (which the sellers in this model are doing), one would like to have somewhat more extreme observations of $r_{i,t}$ so as to obtain better information on its coefficient $\beta_{i,t}$.[13] And since the explanatory variable of the regression also happens to be the control of the system, this can actually be achieved: the experimenting seller realizes that he is like an econometrician who has the luxury of being able to generate his own data points.

[12] In this respect, active learning also has an interesting relationship with Sims' (2003) idea of rational inattention: while both concepts endogenize the signal-to-noise ratio, rational inattention achieves this by endogenizing the amount of *noise* in the system, while active learning on the other hand endogenizes the *signal*.

[13] To see this, note from (5) and (6) that one would not learn anything about $\beta_{i,t}$ if $r_{i,t}$ would remain constant at zero over time. Equation (8) then implies that uncertainty about the true value of $\beta_{i,t}$ will grow with σ^{2}_{η} (the variance of the slope innovation) every period. The reason is that posting the average price is not informative on the perceived degree of product differentiation. More formally, when $r_{i,t}=0$, $\beta_{i,t}$ ceases to play a role in the demand curve (2) - thereby introducing an identification problem with respect to this parameter.

3.4 Solving the model

The seller's problem is given by (1) subject to equations (2)-(8). As shown by Kiefer and Nyarko (1989), this type of problem still admits a contraction mapping argument. In particular, the problem satisfies Blackwell's sufficiency conditions of monotonicity and discounting, as a result of which it has a fixed point (being the value function). Consequently, it can be solved by dynamic programming. The problem has three state variables ($\left\{q_{i,t-1}, b_{i,t|t-1}, \Sigma^b_{i,t|t-1}\right\} \equiv \{q_{i,t-1}, \Xi_{i,t-1}\}$) and the Bellman equation can be written as:

$$V\left(q_{i,t-1}, \Xi_{i,t-1}\right) = \max_{r_{i,t}} \left[\begin{array}{c} \Pi\left(q_{i,t-1}, \Xi_{i,t-1}, r_{i,t}\right) + \delta \int V\left(q_{i,t}, \Xi_{i,t}\right) \\ \times f\left(q_{i,t} | q_{i,t-1}, \Xi_{i,t-1}, r_{i,t}\right) dq \end{array} \right]$$

This equation nicely shows the trade-off between estimation and control that firms face: on the one hand, they want to maximize current period profits $\Pi_{i,t}$, but on the other hand they also have to take the expected continuation value into account (the second term). The latter contains the potential improvements in future performance that may result from past price experimentation and is therefore increasing in the absolute value of $r_{i,t}$. The fact that price experimentation is costly in terms of current period profits, prevents the seller from posting prices that are "too extreme" (even though such actions could reveal the true value of $\beta_{i,t}$). Hereby, the profit maximizing objective of sellers imposes an *endogenous* bound on the precision of the signals they choose to receive. Hence, in this paper's setup, optimizing behavior fulfills a similar role as the exogenous information processing constraint does in rational inattention models.

The model is solved by iterating over the Bellman equation until the value function converges. The exact algorithm is described in Beck and Wieland (2002). It implies that, starting from an initial guess of the value function, one updates the latter by maximizing the right-hand side of the Bellman equation. There, the conditional expectation is evaluated using Gauss-Hermite nodes (since the dependent variable q is normally distributed). To speed up convergence, policy iterations are carried out after each value iteration.

Since these type of control problems are typically characterized by multiple local maxima (see Amman and Kendrick (1995) and footnote 17 of this paper), the algorithm starts with a rough grid search, after which a golden section search is carried out to compute the maximum more precisely.

3.5 Price setting under active learning

We are now in the position to solve the problem set out in Section 3, and see what the introduction of active learning implies for the price setting behavior of firms. In order to do so, Section 5.1 first discusses the model's calibration, after which Section 5.2 describes the model outcomes.

3.5.1 Calibration

As this paper deals with higher-frequency price movements, I calibrate the model at a weekly frequency. Table 1 summarizes the calibration. The weekly discount factor δ is set to imply a standard value for the quarterly discount factor of 0.99. The parameter capturing good-specific habits (γ) equals the estimate of Ravn, Schmitt-Grohé and Uribe (2006) converted to a weekly frequency. The marginal cost of production c is normalized to zero.

I will simplify the analysis by looking at the partial equilibrium case in which the aggregate price level $\overline{p}_t$ is exogenous. In particular, I will just set it equal to a constant, hence $\overline{p}_t = \overline{p} \;\; \forall t$ (the same assumption is effectively made in Matějka (2010a)).

The actual value of $\overline{p}$, along with the values of the other two parameters that determine the location of the demand curve (the average intercept $\overline{\alpha}$ and average slope $\overline{\beta}$), are however difficult to calibrate as their values cannot be estimated that easily. Consequently, these parameters are set in order to ensure two things: i) that sales q_t do not go negative in simulations and ii) that the average price a firm posts over time approximately equals $\overline{p}$ (so that individual and aggregate prices are internally consistent in the model). I have found values of $\overline{\alpha} = 0.38$, $\overline{\beta} = 6$ and $\overline{p} = -1.8$ to work well for these purposes. Since this is a partial equilibrium model, these values do not have direct economic interpretations (they only determine the location of the demand curve), so the negative value for the price level has no meaning here.

Whereas the actual value picked for $\overline{\beta}$ is not that important, the relative uncertainty that $\overline{\beta}$ is surrounded with, is key. The reason is that this determines the seller's experimentation incentives. This relative uncertainty can for example be measured by the t-statistic on β_t. The latter has recently been estimated in the marketing literature by Gorden, Goldfarb and Li (2011). Using household panel data running from January 2001 up to December 2006, they come up with an average t-statistic of about 6 across all 19 grocery categories in their study. This implies that β should get a unit standard deviation, which requires σ_η^2 to equal 0.0199.

Symbol	Interpretation	Value
$\overline{\alpha}$	Average aggregate demand	0.38
$\overline{\beta}$	Average slope of demand curve	6
c	Marginal production cost	0
γ	Habit formation coefficient	0.9875
δ	Discount factor	0.9992
$\overline{p}$	Aggregate price level	−1.8
ρ_β	AR-component in β-process	0.99
σ_μ^2	Variance of intercept innovation	0.0164
σ_η^2	Variance of slope innovation	0.0199

Table 3.1: Calibration.

The autoregressive parameter for $\beta_{i,t}$ is based upon Smets and Wouters (2007), who estimate their equivalent of ρ_β to equal 0.89 on quarterly US data. This leads to a weekly value of 0.99.

Finally, the variance of aggregate demand α (σ_μ^2) is set equal to the variance of the natural log of linearly detrended US nominal GDP from 1947q1 to 2011q1 (taken from the St. Louis Fed).

3.5.2 Model results

To see what each separate step contributes to the final solution, I will start by reducing the problem described by equations (1)-(8), to two simpler problems.

First, one can derive the policy rule for the standard case in which firms live in a perfect information world where they can observe the true values of α_t and $\beta_{i,t}$ without error. In that case, the problem under consideration reduces to a simpler one and one can show that the profit maximizing price equals:

$$p_{i,t}^* = \overline{p} + \frac{1}{2\beta_{i,t}}\left[\alpha_t + \beta_{i,t}\left(c - \overline{p}\right) + \gamma q_{i,t-1}\right] - \delta\gamma \mathbb{E}_t\left\{\frac{q_{i,t+1}}{2\beta_{i,t+1}}\right\} \tag{3.9}$$

Here, both the forward looking term ($\gamma\mathbb{E}_t\left\{q_{i,t+1}/2\beta_{i,t+1}\right\}$) as well as the lagged term ($\gamma q_{i,t-1}/2\beta_{i,t}$) result from the dynamic link that is established by the formation of good-specific habits. In particular, this equation shows the trade-off (described in Klemperer (1995)) sellers face between setting a low price to lock consumers into their specific product (which is what the forward looking term captures) and setting a high price to harvest profits by exploiting consumers who are already "addicted" due to past consumption (this is what is expressed by the lagged term).

Next, we can move on to the specification in which α_t and $\beta_{i,t}$ cannot be observed, as a result of which agents have to learn their true values. In that case, firms could follow the so-called "passive" policy rule. It is however key to note that this rule is not optimal as it implies that the seller does not realize that he is learning from self-generated observations. Consequently, he disregards the link between current and future beliefs (*i.e.* in the optimization step he neglects equations (5)-(8)), as a result of which this learning rule lacks any experimentation motives: although passive learners do update their beliefs as more observations arrive over time, they do not actively seek for better information through price experimentation.

For this case, the passive price setting rule can be shown to equal:[14]

$$p_{i,t}^{\Diamond}=\overline{p}+\frac{1}{2b_{i,t|t-1}}\left[\overline{\alpha}+b_{i,t|t-1}\left(c-\overline{p}\right)+\gamma q_{i,t-1}\right]-\delta\gamma\mathbb{E}_t\left\{\frac{q_{i,t+1}}{2b_{i,t+1|t}}\right\} \tag{3.10}$$

Deviations of the active learning rule from this passive one, measure the extent of price experimentation.

Unfortunately, it is not possible to obtain an analytical expression for the price setting rule under active learning in the model under consideration. We can however characterize an important characteristic of it, namely that the optimal action under active learning will always be more extreme than the one under the passive policy rule. This is done in Proposition 1, which builds upon the following two lemmas. The proofs of these statements can be found in the Appendix.

[14] Note that the passive policy rule is equal to the certainty equivalent one (*i.e.* the rule that one obtains if one replaces actual values by estimates in equation (9)). Also see Balvers and Cosimano (1990, p. 887) on this.

Lemma 1 *Uncertainy on β is decreasing in the squared deviation of the control variable from zero, that is:* $\partial\Sigma^b_{i,t|t}/\partial\left(r^2_{i,t}\right) < 0.$

Lemma 1 formally establishes the previously stated notion that the presence of more extreme observations of the explanatory variable, facilitates the learning process on the unknown coefficient β_t. Lemma 2 shows that this information is actually valuable.[15]

Lemma 2 *The value of the firm is non-increasing in the uncertainty on β, that is:* $\Delta V\left(q_{i,t-1}, b_{i,t|t-1}, \Sigma^b_{i,t|t-1}\right)/\Delta\Sigma^b_{i,t|t-1} \leq 0.$

With help of these two lemmas, we can prove Proposition 1:

Proposition 1 *The relative price set under active learning (call it $r^{\dagger}_{i,t}$) is always more extreme than the price chosen under the passive learning strategy, i.e.* $\left|r^{\dagger}_{i,t}\right| > \left|r^{\diamondsuit}_{i,t}\right|$.

To gain more insight in the effects of active learning, we can numerically solve for the optimal policy rule using the Beck-Wieland algorithm set out in Section 4 of this paper. This was done in order to generate Figure 2. That figure plots the active learning pricing rule and compares it with the passive policy for different combinations of beliefs $\left(b, \Sigma^b\right)$.[16]

The figure clearly shows the experimentation motive: while the passive price setting rule is just a linear, increasing function of demand, the rule resulting from the active learning process contains a kink. This non-linearity shows up in that part of the state space where posting the less extreme, passive learning price would be uninformative on the perceived degree of product differentiation, $\beta_{i,t}$. Hence, the discontinuity simply serves to avoid uninformative actions (that is: it avoids posting a relative price $r_{i,t}$ close to zero).[17]

[15] Note that this lemma does not require differentiability of the value function. The reason is that the latter is not always guaranteed in active learning problems (*cf.* Balvers and Cosimano (1994, upon which the proof of this lemma (and that of Proposition 1) is inspired)).

[16] Note that the perfect information price setting rule (9) equals the passive learning rule (10) for $b_{i,t|t-1} = \beta_{i,t}$ (this is the certainty equivalence property). Consequently, the passive learning and perfect information rules would coincide in Figure 2 as a result of which I have not included the latter explicitly in the figure.

[17] More technically, the reason lies in the fact that the value function is twin-peaked. This is due to the fact that it is a combination of a concave function in $r_{i,t}$ (static profits) and a convex one (the continuation value). The continuation value is increasing in the absolute value of $r_{i,t}$ as

The fact that the policy functions tend to flatten out away from the discontinuity, shows that there is less price experimentation when demand for variety i is either relatively high or low. The reason is that in these extremer parts of the state space the relative price picked by the passive policy (which is the optimal one if one neglects the learning motive) has already moved away from zero, due to the fact that demand is so high/low. Consequently, posting the passive price is not that bad anymore from a learning perspective, thereby reducing the need for additional price experimentation (which is costly as it lowers current period profits). As a result, the active learning rule converges towards the passive one as conditions get more extreme.

Therefore, the optimal price setting rule is shaped like a step function, and prices tend to jump between two rather rigid values: when demand is low, firms post a low price, while an increase in demand above a certain threshold makes them post the higher price. This can be seen as an endogenous form of "local" price stickiness: as long as the model dynamics do not push a seller from the high to the low price platform of his policy function (or vice versa), he is relatively unresponsive to shocks.

Consequently, the experimentation motive is able to replicate the empirical observation (displayed in Figure 1) that individual prices follow a rather discrete pattern over time. This is illustrated in the upper panel of Figure 3, which contains a simulated series (spanning a period of approximately two years) for the control variable $r_{i,t}$ under active learning.[18] Since the optimal price set under active learning varies a bit with beliefs $\left(b_{i,t|t-1}, \Sigma^b_{i,t|t-1}\right)$, and since the upper and lower platforms of the price setting rules are not entirely flat, the resulting price pattern is not exactly discrete. However, it looks very much like it. The reason is that the policy function under active learning actually has a rather flat slope in all dimensions of the problem. With respect to the $b_{i,t|t-1}$-dimension, the intuition is similar to that for the $q_{i,t-1}$-dimension: less conscious price experimentation is needed when conditions are more extreme (*i.e.* when $\beta_{i,t}$ is believed to be either

more extreme relative prices are more informative. This generates two local maxima (one where $r_{i,t} < 0$ and one where $r_{i,t} > 0$), from which the max-operator in the Bellman equation selects the one yielding the highest value. This leads to an avoidance of the region where the continuation value is minimized, *i.e.* that region where $r_{i,t}$ is "too close" to zero.

[18] This particular figure was constructed by giving the seller correct initial beliefs about the true value of β_1, but with an initial uncertainty $\Sigma^b_{1|0} = 0.595$. The typical pattern arising in simulations is robust to choosing different initial beliefs.

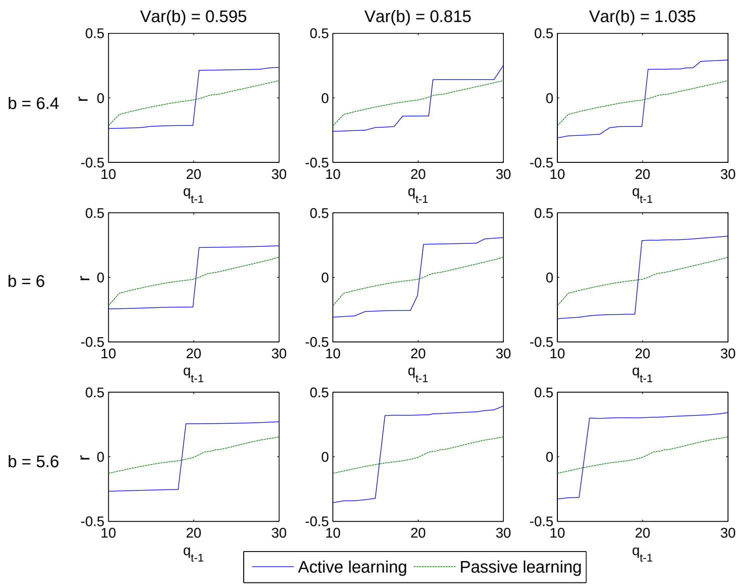

Figure 3.2: Comparison of active and passive price setting rules for different beliefs.

very high or very low). In those circumstances the passive learner already posts a relative price that has moved away from zero (thereby making it more informative) and the active learner does not have to undertake so much costly action on top. With respect to $\Sigma^b_{i,t|t-1}$, the fact that the slope parameter is time-varying (which is motivated by evidence from the marketing literature; *cf.* Gordon, Goldfarb and Li (2011)) implies that uncertainty continues to renew itself and that the learning process never stops. Consequently, uncertainty on the estimate of $\beta_{i,t}$ remains relatively stable over time and the slope in that dimension only gets to play a subordinate role.

Adding a small menu cost would probably increase the degree of discreteness in Figure 3 as the small price changes that can be observed in the current simulation would then no longer be worthwhile (a "region of inaction" would arise). Moreover, a menu cost could also be helpful in decreasing the frequency of price changes: in the model, prices tend to change from high to low (or vice versa) about once every two weeks, which is slightly more often than (but close to) the findings of Kehoe and Midrigan (2010) and Eichenbaum, Jaimovich and Rebelo (2011) (both studies report that posted prices tend to change about once every three weeks in their data sets).

The price path arising under active learning bears similarity to the one obtained by Matějka (2010a). Sellers in his model have limited information capacity, as a result of which it becomes optimal for them to pay attention to a source of information that provides a small number of different signals only. Such a signal could be the first digit of unit input costs and since a digit can only take a few different values, there are only a few different signals in his framework - leading to the discreteness.[19]

In the present paper, the (near-)discreteness arises purely from the active learning process: apart from the latter, the model is a rather standard profit maximization problem without price adjustment costs. This can also be seen by looking at the two lower panels of Figure 3. Those show the price paths in response to the same series of shocks under both passive learning as well as under perfect information (the case in which sellers can simply observe the true values of α_t and $\beta_{i,t}$). As one can see, these series no longer show any discreteness (a result of the fact that the passive and perfect information policy rules are not kinked).

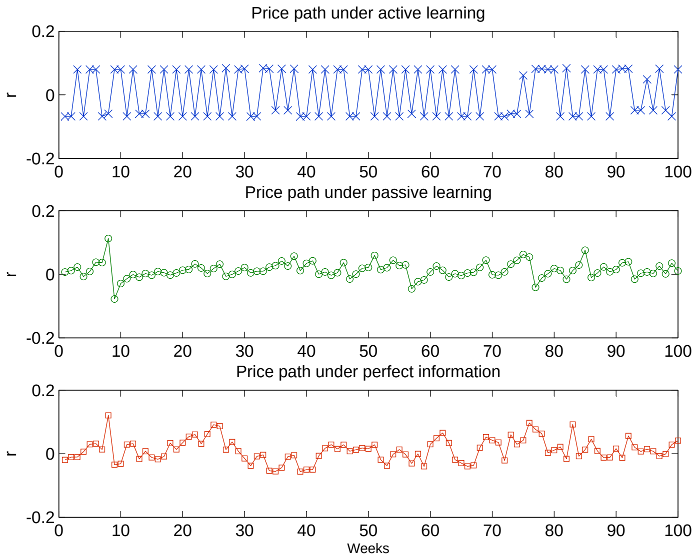

Figure 3.3: Simulated paths for the relative price posted under active learning, passive learning and under perfect information.

Also note that passively learning sellers have no problems with posting a rel-

[19] In Matějka (2010b), the discreteness follows from modeling the consumer as being rationally inattentive.

ative price that is close to zero, as they do not realize that this obstructs their learning process on the degree of product differentiation, $\beta_{i,t}$. Sellers who are learning in the optimal active manner do realize this, as a result of which they avoid posting a zero relative price (this is what the kink in the policy function takes care of). Consequently, learning is faster in the active specification. This is illustrated by Figure 4, which was generated by giving a seller incorrect initial beliefs on the value of β (namely that its value equals 7 instead of 6), after which both an active and a passive learner start updating their beliefs. During this learning process, all shocks were shut down to illustrate the process of convergence towards the truth. As can be seen from the figure, learning is much faster when it is done in an active manner, since an active learner consciously generates more extreme (and thereby more informative) observations.

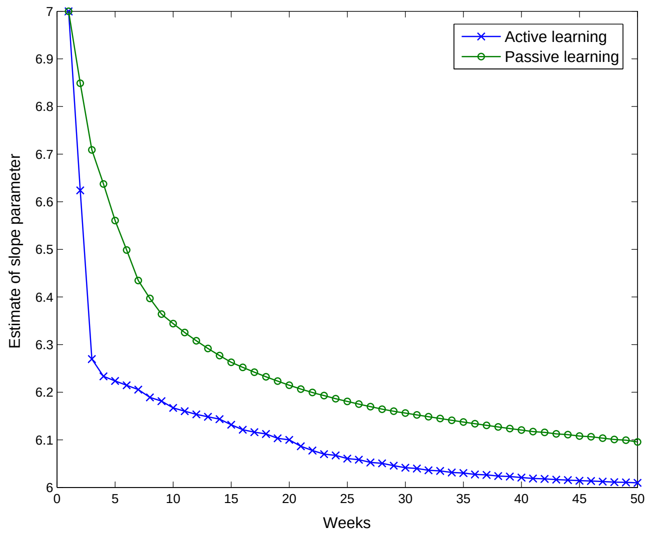

Figure 3.4: Learning process on β under both active and passive learning. Initial estimate is 7, while the true value of β equals 6.

Turning back to Figure 3, one can see that the model is also able to replicate the empirical observation that sellers *alternate* between posting the high and the low price. This would not happen if there were no good-specific habits. The reason is that, from an informational point of view, a low price is just as informative as a high one with the same distance from the aggregate price level. As noted by Bachmann and Moscarini (2011), what matters is the absolute deviation of $p_{i,t}$ from $\overline{p}$, not its sign. In the most standard setting sellers would therefore just post the high (or low) price, and stick to that choice ever after (this happens in Bachmann and Moscarini (2011)). But with good-specific habits, the demand curve contains a lagged dependent variable which introduces an additional dynamic link: a seller who posts an above average price in period t (*i.e.* someone who has set $r_{i,t} > 0$) will sell less units in that period (*cf.* the demand curve (2)) as a result of which he faces a lower endogenous state variable (the stock of habits $q_{i,t}$) in the next period. Consequently, he is pushed towards the lower platform of his policy function at time $t+1$, where it is optimal to post the lower price.

Figure 3 also illustrates that the experimentation motive introduces additional volatility in the price series (driven by the combination of the seller's preference for more extreme observations, with the aforementioned dynamics due to good-specific habits). In this sense, it may be an explanation for the observation that prices are more volatile than underlying marginal costs (as documented by Eichenbaum, Jaimovich and Rebelo (2011) and Carlsson and Skans (2011)). Time-variation in the slope of the demand curve already helps (due to this, prices under perfect information and passive learning are also more volatile than marginal costs (which are constant in this paper)), but the experimentation motive gives this channel an extra kick.

Related to this, it also links to Golosov and Lucas (2007) (and the large literature following that paper), who argue that the relatively large idiosyncratic price movements observed in the data indicate the presence of large idiosyncratic shocks. This paper however shows that it do not necessarily have to be large shocks driving these movements; instead, it could be that much smaller shocks are being amplified by the endogenous, idiosyncratic experimentation process of each individual seller. In this respect, De Graeve and Walentin (2011) actually provide evidence that idiosyncratic shocks are not as large as previously thought (thereby questioning the Golosov-Lucas calibration), while Nakamura (2008) argues that the large shocks to manufacturers' productivity assumed by Golosov and Lucas (2007) would generate substantial comovement across prices for the same good at different stores, which is not observed in the data.

Finally, one can analyze how the aggregate price level in the active learning model responds to aggregate shocks, and compare this response to the ones obtained under passive learning and perfect information. Following Matějka (2010a), I do this by simulating, and averaging over, the behavior of a large number of agents. Although this paper's partial equilibrium setup is not ideal to study this issue (since an aggregate shock will change the aggregate price level $\overline{p}$, which is assumed to be constant in the model), I believe it does capture the core of the underlying learning dynamics that play a role here.[20]

To tackle this question, we need to make a slight modification to the model considered so far. In particular, we need to introduce aggregate risk with respect to the slope of the demand curve. In the benchmark model, aggregate conditions were captured by the common mean value of the process for β (indicated by $\overline{\beta}$). There, this mean was however assumed to be known and kept constant over time, which is going to be relaxed in this section.

The most insightful way to do this, and simultaneously illustrate the accompanying learning dynamics that are at play, is to assume that the economy can be in two regimes: a boom or a recession. Associated with these two regimes, are two values for the long run average of the slope parameter: $\overline{\beta}_H$ (applicable in a recession) and $\overline{\beta}_L$ (applicable in a boom). I will assume that consumers are more price sensitive in recessions (*i.e.* $\overline{\beta}_H > \overline{\beta}_L$), which is intuitive and also line with empirical findings of Gordon, Goldfarb and Li (2011). The actual realisations of $\overline{\beta}_t$ will however remain unobserved to the sellers in the model (making it a two-state hidden Markov problem), so they have to learn which regime the economy is in.

Related to this learning process, the two-state Markov assumption implies that beliefs are conveniently characterized by only one additional state variable: $\lambda_{i,t|t-1} = \Pr_i(\overline{\beta}_t = \overline{\beta}_H)$, which is seller i's prior time t belief that the average price elasticity is high. This limits the dimensionality curse associated with the solution method, while still capturing the notion of aggregate risk (see Ellison and Valla (2001) for a similar approach).

The cost of introducing Markov switching is that it is no longer possible to assume that idiosyncratic shocks $\eta_{i,t}$ enter a persistent $AR(1)$-process. The reason is that Bayesian updating in a mixed $AR(1)$/Markov switching model no longer admits an analytical solution. Instead, simulation methods such as the Gibbs-sampler must be employed (*cf.* Kim and Nelson (1999)), which is not feasible

[20] In fact, if anything I believe that allowing $\overline{p}$ to change following aggregate shocks would only strengthen the conclusions that are to follow if $\overline{p}$ is unobserved as well. The reason is that sellers then have to learn about an extra parameter (and potentially form higher order beliefs as in Nimark (2008)), which complicates their learning process - thereby reinforcing the upcoming results (which are driven by the sluggishness stemming from the learning process).

within this model's framework.

Consequently, I proceed by assuming that all persistence comes through the mean value for β. Since idiosyncatic movements in the slope parameter would cancel out in the aggregate (irrespective of the issue whether they are persistent or not), and since this specification still contains an active learning motive with respect to the slope term, I believe that this assumption does not seriously limit the generality of the exercise that is to follow.

With respect to the persistence of the mean value for β, the parameter π expresses the conditional probability of staying in the current regime (so $\pi = \Pr(\overline{\beta}_{t+1} = \overline{\beta}_J | \overline{\beta}_t = \overline{\beta}_J)$, $J \in \{H, L\}$). In the simulation I set $\pi = 0.998$ (as in Matějka (2010a)), which implies that there is a 10 percent probability that the aggregate state changes in any given year.

Summarizing, the process for the slope parameter can be described by:

$$\beta_{i,t} = \overline{\beta}_t + \eta_{i,t}, \text{ with } \eta_{i,t} \sim N(0, \sigma_\eta^2) \tag{3.11}$$

$$\overline{\beta}_t \in \{\overline{\beta}_H, \overline{\beta}_L\} \tag{3.12}$$

By working on Bayes' rule, one can show that the belief updating process now depends on the relative probability of observing the actual signal in either of the two regimes:

$$\lambda_{i,t|t} = \frac{\lambda_{i,t|t-1} P(q_{i,t}|\overline{\beta}_t = \overline{\beta}_H)}{\lambda_{i,t|t-1} P(q_{i,t}|\overline{\beta}_t = \overline{\beta}_H) + \left(1 - \lambda_{i,t|t-1}\right) P(q_{i,t}|\overline{\beta}_t = \overline{\beta}_L)}, \tag{3.13}$$

with $P(q_{i,t}|\overline{\beta}_t = \overline{\beta}_J) = \exp\left(-0.5\left[q_{i,t} - \overline{\alpha} + \beta_J r_{i,t} - \gamma q_{i,t-1}\right]^2 / \sigma_\mu^2\right)$, $J \in \{H, L\}$.

Since there is a $(1-\pi)$ probability that the regime switches in going from the end of period t to the beginning of period $t+1$, we have that:

$$\lambda_{i,t+1|t} = \pi \lambda_{i,t|t} + (1-\pi)\left(1 - \lambda_{i,t|t}\right) \tag{3.14}$$

Compared to Section 3, the other elements of the setup (as well as the calibration) are left unchanged so I do not repeat them here.

The results of this exercise are shown in Figure 5. That figure shows the posted price averaged over 5,000 sellers. During the simulation, all individual sellers were hit by two types of disturbances: idiosyncratic shocks (all sellers have their individual series for $\eta_{i,t}$, for example representing the popularity of the particular brand they are selling) and aggregate ones (all these individual series for $\beta_{i,t}$ evolve around a common mean slope of $\overline{\beta}$, capturing aggregate economic conditions). For the first 9 weeks, the common mean value of $\beta_{i,t}$ is set equal to 6.5. Then, in the 10th week, a positive aggregate shock to the economy suddenly makes all consumers less price sensitive, as a result of which $\overline{\beta}$ falls to 5.5.[21]

Several things are to be noted from the figure. First observe that although individual prices show a lot of variation due to the experimentation motive (recall Figure 3), the aggregate price level evolves smoothly in the absence of aggregate shocks. The reason is that all experimentation is idiosyncratic to each individual seller, as a result of which it cancels out in the aggregate. Although this is probably an exaggeration of the amount of idiosyncrasy observed in reality, the dataset analyzed in Nakamura (2008) does hint in this direction, as she reports that only 16 percent of all price variation is common to different stores selling an identical product.

However, when the aggregate shock hits in period 10, the price level does respond in all three specifications. In particular, it goes up, which reflects the increase in market power for sellers associated with the decrease in the average price sensitivity of consumers. Next to that, all three models also produce a hump-shaped response to the shock. This is due to the existence of good-specific habits: consumers are "addicted" to the specific good they have been consuming in the past, as a result of which sellers temporarily obtain the power to charge prices that lie above their new long run values.

The persistence of this hump shape however differs between the three models. In particular, the fact that the squared line in the upper panel of Figure 4 converges rather quickly to its new value shows that habit formation in itself (that is: without the help of informational imperfections) does not manage to generate much persistence. As the circled response shows, assuming that the true values of $\beta_{i,t}$ and $\overline{\beta}_t$ cannot be observed, helps in this respect. The reason is that the learning process implies that firms only gradually find out that the mean value of the slope parameter has changed. Because a regime change is not that likely

[21] The size of the imposed shock probably is unrealistically large (especially given the fact that it is assumed to take place within a week), but considering a more gradual, persistent change from $\overline{\beta}_H$ to $\overline{\beta}_L$ would only increase the sluggishness in the responses of aggregate prices - thereby strengthening the findings that emerge from Figure 5.

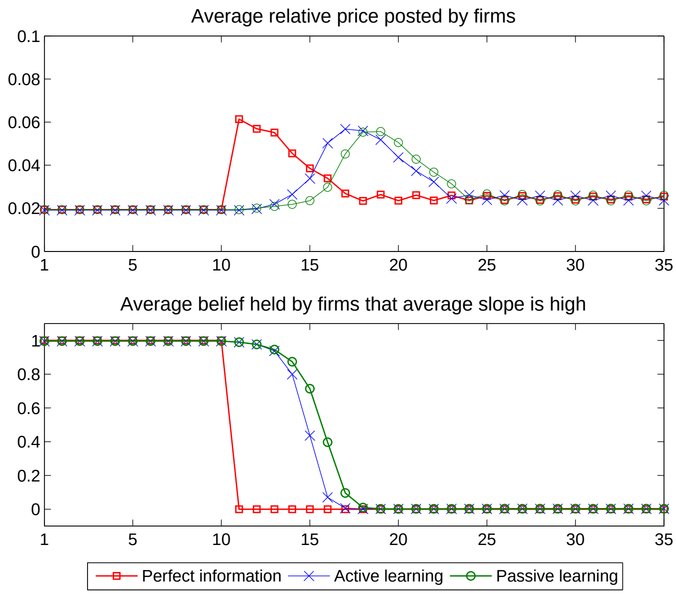

Figure 3.5: Average posted price (top panel) and average beliefs (bottom panel) under active learning, passive learning and under perfect information.

a priori, sellers initially attribute the "anomalous" observations they receive to transitory factors to which they do not wish to respond. Only as they continue to receive observations that are difficult to reconcile with the old regime, they gradually become convinced that a regime switch has occurred.

The circled response was however generated under the assumption that learning takes place in the suboptimal passive manner. That is: firms post the certainty equivalent price and update their beliefs with whatever information this pricing strategy happens to generate.

Under active learning in contrast, firms internalize this learning process and consciously use their pricing strategy to this end. Consequently, learning is faster compared to the passive case (see the lower panel of Figure 5), as a result of which the aggregate shock is propagated to a somewhat lesser extent. However, average beliefs still need time to converge to the new value of $\overline{\beta}$, so the aggregate price level continues to move in an inertial way. Hence, one can conclude that the active leaning process of sellers who are trying to figure out the true value of their slope parameter, is not only able to replicate the volatile, discrete pattern followed by individual prices, but is also consistent with the idea that the aggregate price level responds sluggishly to aggregate shocks. In this respect, we can thus conclude that individual price flexibility does not translate into aggregate price flexibility when individual price changes are driven by the experimentation motive of a seller who is learning the slope of his demand curve.

3.6 Discussion

Active learning problems rapidly become quite challenging to solve. In order to advance, this paper therefore had to abstract from several interesting issues.

First, the analysis in this paper is only partial equilibrium in nature. In this respect, contemporaneous work by Bachmann and Moscarini (2011) forms an important contribution. In order to generate "endogenous uncertainty", they manage to place the experimenting seller in a dynamic stochastic general equilibrium environment. Although this requires them to simplify in some dimensions (they for example reduce the problem to a two-armed bandit setting, where the price elasticity of demand can take on only two values: high or low), the main idea underlying the present paper should carry over to their setup. Investigating whether it is possible to construct a general equilibrium active learning model that is consistent with the behavior of prices at both the micro and macro level, would therefore be a logical next step along this line of research. Subsequently, one could investigate to what extent the experimenting seller model is able to replicate the quantitative aspects of the data.

Second, this paper simplifies by assuming that firms are using their pricing strategy to learn about only one of the parameters of the demand curve. This avoids the dimensionality curse that is inherent to the solution method.[22] In particular, I have assumed that firms are only using their pricing strategy to facilitate the learning process on their slope parameter. The full problem, in which firms are simultaneously learning on both slope and intercept of an equation that has the form of (2), is studied in Wieland (2006) - although in a very different context: he analyzes a central bank that is trying to learn the Phillips curve in

[22] The problem in which sellers are actively learning about both the intercept and slope of the demand curve would have six state variables: $q_{i,t-1}$, two estimated means, two variances and one covariance estimate. At a weekly frequency (which implies a discount factor close to one), the model becomes very difficult to solve. Potentially, such a bigger problem could be tackled with faster algorithms that approximate the original problem (such as the one in Cosimano (2008)).

an active manner. There he actually finds that the extent of experimentation is *greater* if learning takes place with respect to both intercept and slope, rather than just with respect to the slope (although the differences are small).[23] This suggests that the results obtained in the present paper are robust to the more general setting in which firms are learning simultaneously on multiple parameters of the demand curve, although this of course remains to be verified for the particular problem under consideration.

Third, I see this paper's sale motive as only part of the complete story, since generating information on demand is unlikely to be the only reason for firms to hold sales. It is however important to note that the experimentation motive is fully complementary to other sale motives offered by the literature (such as Matějka (2010a,b), Guimarães and Sheedy (2011) and Nakamura and Steinsson (2011)), so the channel emphasized by this paper is by no means at odds with these earlier contributions. In addition, this paper abstracts from potential forces at work in reality that make firms want to abstain from price changes. Menu costs are an obvious example, while there may also exist information-related reasons for this (see *e.g.* L'Huillier (2011) where sellers want to delay price adjustments to keep consumers uninformed).

Fourth, consumers behave rather passively in this paper (the consumer side is simply represented by the demand curve (2)). Consequently, consumers do not take into account that sellers have an experimentation motive. If consumers would be aware of this, strategic interactions between the two sides of the market might arise - the analysis of which I leave for future work.

Fifth, I do not consider changes to marginal costs. Eichenbaum, Jaimovich and Rebelo (2011) suggest that the latter are important, as price changes very often coincide with cost changes in their data set. Although there does not seem to be a consensus view on this issue yet,[24] adding cost shocks to the model could be an interesting next step. This modification would not eliminate the experimentation motive with respect to the slope term, but if costs are imperfectly observed as well,

[23] Because Wieland (2006) is not interested in the higher frequency features of the data, he can calibrate his model at a yearly frequency. This allows him to set the discount factor equal to 0.95, which greatly facilitates the convergence of the value function. Nevertheless, he reports that solving his model still takes 3 days on a 2.21 GHz AMD processor with 1.48 GB of RAM.

[24] Nakamura's (2008) results for example suggest that cost shocks only play a minor role in explaining high-frequency price movements; similarly, Chevalier and Kashyap (2011) document that prices in Dominick's data set are also often adjusted without movements in costs.

it does give the seller an incentive to produce more units as that would improve the precision of his cost signal (see Gal-Or (1988)). In addition, changes in marginal costs might also produce shifts in the pricing band, just as we observe in the data where the high and/or low price occasionally settle at new levels.

Finally, since both this paper's approach (recall footnote 12) as well as its results are rather similar to that of certain papers in the rational inattention literature, it may be worth pointing out some interesting differences. First, the model developed in this paper does not predict a sharp distinction between a seller's response to aggregate and idiosyncratic shocks. This contrasts with Maćkowiak and Wiederholt (2009), where sellers allocate most of their attention to idiosyncratic conditions. As a result, sellers in their model quickly react to changes in the latter, while responding rather sluggishly to aggregate developments. Although Boivin, Giannoni and Mihov (2009) and Maćkowiak, Moench and Wiederholt (2009) present evidence that this indeed is the case in reality, the recent papers by De Graeve and Walentin (2011) and Carlsson and Skans (2011) question this. They find that firms respond sluggishly to *both* aggregate *and* idiosyncratic conditions - a finding that is more in line with a learning model in which sellers face uncertainty on the persistence of both shock types. Hopefully, future work can shed more light on these contrasting views.

Second, compared to the rational inattention approach, active learning imposes somewhat more discipline on the model builder: since active learners behave in a fully rational way, the model does not contain a free parameter whose value is difficult to calibrate (like the information processing constraint in rational inattention models). Instead, in this paper's setup, the costs of more precise signals are determined *endogenously* by the amount of profits a seller has to give up to experiment. However, despite the absence of such a degree of freedom, the model is still able to produce rather similar results as the rational inattention models in this literature.

3.7 Conclusion

The standard model makes the rather strong assumption that sellers can observe their entire demand curve. This paper relaxes this assumption by making the demand curve unobserved and analyzes the optimal pricing strategy of a seller that operates in such an environment. It is shown that when learning is modeled in an optimal way, one can not only match the behavior of aggregate variables (a point that has been made before already), but one is also able to replicate certain micro-dimensions of the data (a finding that is new to the best of my knowledge). In particular, as achieved by Matějka (2010a,b), the model constructed in this paper manages to generate all of the following three stylized facts that seem present in reality:

1. Individual prices change frequently.
2. Individual prices tend to move back and forth between only a few rigid values.
3. The aggregate price level responds gradually to aggregate shocks.

Note that the model's ability to replicate these facts is *not* due to some form of irrationality: all results are driven by the rational behavior of sellers who are responding optimally to the fact that they cannot observe their entire demand curve. Given that more and more retailers are using sophisticated price optimization software (of which this paper's model is essentially a mini-version; also recall footnote 3), rational price setting can not only be thought of as an important optimal benchmark that approximates the real economy and is worth analyzing (like rational expectations), but also as actually materializing to a first degree in practice.[25]

The model can match stylized fact #1 because the introduction of active learning gives firms an incentive to experiment with their price. Next to the fact that price experimentation is optimal for sellers facing uncertainty on the slope of their demand curve, there is also empirical evidence that this type of behavior plays a role in reality (see Pashigian (1988, who documented a positive impact of demand uncertainty on the frequency of sales), Campbell and Eden (2010, who analyzed US scanner data), Einav *et al.* (2011, who looked at eBay-listings) and Gaur and Fisher (2005, who held a survey among large US retailers)). Experimentation could also be the reason why prices are more volatile than marginal costs in the

[25] In this respect, one can also take the idea of this paper one step further and develop a new class of macroeconomic models with *actual* pricing algorithms at its core. Since these algorithms determine prices for such a large share of the economy, it may be interesting to investigate how the prices they set respond to typical economic shocks.

data. Simultaneously, it nuances the point of view (expressed by Golosov and Lucas (2007), among many others) that the large idiosyncratic variations in prices observed in the data indicate the presence of large idiosyncratic *shocks*; instead, this paper shows that these large movements could also stem from the *endogenous* experimentation motive of individual sellers.

In addition, the experimenting seller model can go a long way in reproducing the discreteness of individual prices observed in reality (stylized fact #2). Replicating this feature of the data has proved to be a major challenge to most price setting models. Although the discreteness arising in this paper is not exact, it looks very much like it. The reason is that the experimentation motive shapes the price setting rules like a step function: they have a high price platform (associated with high demand) and a low one (associated with low demand), separated by a sharp kink. Consequently, the model displays a form of endogenous, local price stickiness (despite the absence of price adjustment costs): as long as a shock does not move a seller from the high to the low platform (or vice versa), it is optimal for him to keep his price rather fixed - even if demand varies locally.

Finally, notwithstanding the flexibility of individual prices under active learning, the aggregate price level still responds sluggishly to shocks. This is due to the fact that average beliefs change only gradually over time. Consequently, the model is also able to capture the inertia found in prices at the aggregate level (stylized fact #3).[26] To the extent that price changes at the individual level are driven by the experimentation motive of a seller who is actively learning the slope of his demand curve, the flexibility of prices at the micro stage thus does not translate into aggregate price level flexibility.

[26] In conversations, I always liked to simplify the message of this paper as follows: this paper presents a model of "active learning". The "active"-part of the setup delivers the volatility and (near-)discreteness of individual prices, while the fact that there is still some "learning" going on, generates the sluggish responses of prices to aggregate shocks.

3.8 Appendix

Lemma 1 *Uncertainy on β is decreasing in the squared deviation of the control variable from zero, that is:* $\partial\Sigma^b_{i,t|t}/\partial\left(r^2_{i,t}\right) < 0$.

Proof. First define:

$$u \equiv \Sigma^b_{i,t|t-1} - \frac{\left(\Sigma^b_{i,t|t-1}\right)^2 r^2_{i,t}}{r^2_{i,t}\Sigma^b_{i,t|t-1} + \sigma^2_\mu}$$
$$v \equiv r^2_{i,t}$$

Differentiating these with respect to $r_{i,t}$ yields:

$$\frac{\partial u}{\partial r_{i,t}} = -\frac{2r_{i,t}\left(\Sigma^b_{i,t|t-1}\right)^2\sigma^2_\mu}{\left(r^2_{i,t}\Sigma^b_{i,t|t-1} + \sigma^2_\mu\right)^2}$$
$$\frac{\partial v}{\partial r_{i,t}} = 2r_{i,t}$$

Via the chain rule we then obtain that:

$$\frac{\partial\Sigma^b_{i,t|t}}{\partial\left(r^2_{i,t}\right)} = \frac{\partial u}{\partial v} = \frac{\partial u}{\partial r_{i,t}}\cdot\frac{\partial r_{i,t}}{\partial v} = -\frac{\left(\Sigma^b_{i,t|t-1}\right)^2\sigma^2_\mu}{\left(r^2_{i,t}\Sigma^b_{i,t|t-1} + \sigma^2_\mu\right)^2},$$

which is strictly negative. ■

Lemma 2 *The value of the firm is non-increasing in the uncertainty on β, that is:* $\Delta V\left(q_{i,t-1}, b_{i,t|t-1}, \Sigma^b_{i,t|t-1}\right)/\Delta\Sigma^b_{i,t|t-1} \leq 0$.

Proof. Fix the optimal policy sequence $r_{i,t+j}(\Sigma^b_{i,t|t-1})$ for all $j \geq 0$. Now decrease $\Sigma^b_{i,t|t-1}$ to any $\Sigma^b < \Sigma^b_{i,t|t-1}$ leaving the $r_{i,t+j}$-sequence unchanged. Call the associated (non-optimized) value function $W\left(q_{i,t-1}, b_{i,t|t-1}, \Sigma^b, r(\Sigma^b_{i,t|t-1})\right)$ (where the last argument expresses that the $r_{i,t+j}$-sequence continues to be based upon the old level of uncertainty $\Sigma^b_{i,t|t-1}$). Now compare this value to the optimized value function $V\left(q_{i,t-1}, b_{i,t|t-1}, \Sigma^b_{i,t|t-1}\right)$.

To do so, first observe from the profit function in equation (1) that current period profits are not affected by the change in uncertainty. Furthermore note that the change in uncertainty does not affect future pay-offs either, since the firm continues to set its control based upon the old level of uncertainty $\Sigma^b_{i,t|t-1}$. Consequently, $W\left(q_{i,t-1}, b_{i,t|t-1}, \Sigma^b, r(\Sigma^b_{i,t|t-1})\right) = V\left(q_{i,t-1}, b_{i,t|t-1}, \Sigma^b_{i,t|t-1}\right)$. In addition, optimization implies that $V\left(q_{i,t-1}, b_{i,t|t-1}, \Sigma^b\right) \geq W\left(q_{i,t-1}, b_{i,t|t-1}, \Sigma^b, r(\Sigma^b_{i,t|t-1})\right)$. Putting these two results together tells us that

$$V\left(q_{i,t-1}, b_{i,t|t-1}, \Sigma^b\right) \geq V\left(q_{i,t-1}, b_{i,t|t-1}, \Sigma^b_{i,t|t-1}\right),$$

so firm value is non-increasing in uncertainty on β. ■

Proposition 1 *The relative price set under active learning (call it $r^{\dagger}_{i,t}$) is always more extreme than the price chosen under the passive learning strategy, i.e. $\left|r^{\dagger}_{i,t}\right| > \left|r^{\Diamond}_{i,t}\right|$.*

Proof. First, consider the case for which $r^{\Diamond}_{i,t} > 0$. Take any $r^{\bullet}_{i,t} < r^{\Diamond}_{i,t}$. Note that this cannot be optimal since $r^{\Diamond}_{i,t}$ maximizes current period profits, while $r^{\Diamond}_{i,t}$ is also more valuable for future periods by Lemmas 1 and 2 (since it is more extreme). Hence, $r^{\dagger}_{i,t} \geq r^{\Diamond}_{i,t}$ is required for optimality.

By an envelope theorem argument it then follows that $r^{\dagger}_{i,t} > r^{\Diamond}_{i,t}$. After all, starting from $r^{\Diamond}_{i,t}$, variations in $r_{i,t}$ only affect current profits to a second order (since $r^{\Diamond}_{i,t}$ maximizes current period profits), while increasing $r_{i,t}$ increases the value of the firm to a first order by Lemmas 1 and 2.

The proof for the opposite case in which $r^{\Diamond}_{i,t} < 0$ follows analogously, *mutatis mutandis*. ■